RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Conseil d'évaluation de l'École

Document de travail

Document de travail 22-01 — Février 2022

Renforcer la cohérence des évaluations des acquis des élèves : un enjeu de réussite et d'équité

Conseil d'évaluation de l'École

Directrice de la publication : Béatrice Gille – présidente du CEE

Documents d'appui du Groupe de travail sur l'évaluation des acquis des élèves

Table des matières

Thème 1 : Repères sur les évaluations standardisées en France 2

1. Repères historiques 2
2. Cartographie des évaluations standardisées 3
3. Méthodologie des évaluations standardisées 4
4. Qu'évaluent les évaluations standardisées ? 4

Tableau 1 : Synthèse des évaluations standardisées réalisées en France 5

Tableau 2 : Parcours scolaire d'un élève et évaluations standardisées nationales et examens 8

Thème 2 : Revue de littérature sur l'évaluation des acquis des élèves 9

1. Cadre d'analyse général de l'évaluation des acquis des élèves au sein du système éducatif ... 9
2. Les évaluations standardisées 11

Un recours croissant aux évaluations standardisées 11

Réceptions, usages et effets des évaluations standardisées des acquis des élèves 12

Apports et limites des évaluations internationales 18

3. L'évaluation dans la classe et les examens 19

Les pratiques d'évaluation à l'école primaire 19

Les pratiques d'évaluation dans le second degré 20

Jugement scolaire et biais évaluatifs 23

4. L'évaluation par les examens 25

Bibliographie 27

Liste des sigles utilisés 31

Thème 1 : Repères sur les évaluations standardisées en France

1. Repères historiques

En France, on distingue trois périodes principales dans la mise en œuvre des évaluations standardisées (Rocher et Trosseille, 2015)[1] :

Années 1980 : mise en œuvre à partir de 1979 d'évaluations bilans sur des échantillons représentatifs d'élèves à l'école et au collège. Elles portaient au départ sur le français et les mathématiques avant d'être élargies aux autres disciplines. L'objectif était de disposer d'un véritable « observatoire permanent des acquis des élèves » (Levasseur, 1996). Ainsi, tout au long des années 1980, les différents niveaux de la scolarité seront, année par année, l'objet de ces évaluations.

Années 1990 : la loi d'orientation de 1989 dispose que « la scolarité est organisée en trois cycles pour lesquels sont définis des objectifs et des programmes nationaux de formation comportant une progression annuelle ainsi que des critères d'évaluation ». Afin de fournir aux enseignants les repères nécessaires et les aider à identifier les besoins des élèves, des évaluations diagnostiques sont mises en place pour tous les élèves à l'entrée des classes de CE2, 6ème et 2nde. Ces évaluations sont maintenues en CE2 et 6ème jusqu'en 2007-2008. Sur cette période, les évaluations sur échantillon (type bilan) deviennent plus irrégulières.

Années 2000 : suite à une recommandation du HCéé, mise en œuvre en 2003 du Cycle des évaluations disciplinaires réalisées sur échantillons (Cedre) qui évalue les acquis des élèves de CM2 et 3ème. Des domaines différents sont évalués chaque année. L'un des atouts de ces évaluations est de permettre des comparaisons diachroniques. En 2008, de nouvelles évaluations sont mises en place en CE1 et CM2, avec une double finalité : dresser un bilan des acquis fortement adossé aux programmes et établir un diagnostic des acquis en s'adressant à l'ensemble des élèves. Ces évaluations, fortement contestées au moment de leur mise en place, deviennent facultatives en 2013 puis seront abandonnées en 2014.

À partir de septembre 2017, des évaluations exhaustives sont mises en place tout d'abord en CP-CE1 puis en 6ème et 2nde. Elles portent sur les acquis en français et mathématiques et sont complétées en 2020 par un test de fluence en 6ème.

Les années 2000 sont aussi marquées par la participation de la France à de nombreuses enquêtes internationales dont les résultats sont de plus en plus médiatisés (TIMSS, PIRLS, PISA).

[1] Les références complètes sont présentées dans la bibliographie à la fin de ce document.

2. Cartographie des évaluations standardisées

Un tableau récapitulatif des évaluations standardisées conduites en France est présenté ci- dessous.

Les enquêtes nationales exhaustives (début CP, mi-CP, CE1, 6ème, 2nde) ont une double finalité :

- finalité pédagogique : « identifier les besoins spécifiques de chaque élève, afin d'aider les enseignants à mieux adapter leurs modalités d'intervention » ; aider les professeurs à « déterminer précisément où les compétences de leurs élèves se situent, non seulement au sein de leur classe et de leur établissement, mais également par rapport à tous les élèves français ». Elles sont placées en début d'année scolaire (et à la moitié de l'année en CP) « pour donner aux enseignants le temps de mettre en place des interventions pédagogiques sur toute l'année » (Évaluer pour mieux aider, CSEN) ;
- finalité de pilotage : « permettre à notre École de disposer d'éléments consolidés susceptibles d'éclairer les effets des politiques éducatives » (Communiqué de presse du 9 novembre 2020, MENJS). La synthèse des résultats des élèves donne des outils de pilotage à tous les niveaux (école, circonscription, collège, lycée, département, académie, national).

Ces évaluations portent uniquement sur le français et les mathématiques. En 6ème, elles ont été complétées par un test de fluence en 2020. Ces évaluations sont conçues par le CSEN avec l'appui de la DEPP dont le rôle est important en termes d'ingénierie et d'opérationnalisation. L'IGÉSR et la DGESCO sont également parties prenantes (expertise et conseil). Les résultats sont diffusés lors de conférences de presse (DGESCO, DEPP) et sont publiés par la DEPP (*Notes d'information)* et le CSEN (*Notes du CSEN*)

Les enquêtes nationales sur échantillon permettent d'établir un bilan des acquis des élèves. Elles ont principalement une finalité de pilotage. Les objets des évaluations sont diversifiés et portent sur plusieurs disciplines scolaires, le socle commun et certaines compétences socio-cognitives. Elles sont construites et pilotées par la DEPP, en lien avec l'IGÉSR et la DGESCO. Leurs résultats sont diffusés par la DEPP : conférence de presse, *Notes d'information*, articles dans la revue *Éducation & Formation*.

Les enquêtes internationales sont construites par des organismes internationaux. Les objets d'évaluation concernent des éléments de littéracie (compréhension de l'écrit, culture mathématique, culture scientifique, littéracie numérique). En 2022, la France participera pour la première fois à une enquête testant les « *connaissances et compétences civiques, attitudes et comportements* ». La DEPP, en tant qu'opérateur national, relaie les résultats de ces enquêtes (*Notes d'information*, articles dans la revue *Éducation & Formation*).

Toutes les évaluations standardisées portent sur le temps de la scolarité obligatoire (cf. ci- dessous, tableau 2). Aucune évaluation standardisée ne concerne les élèves au lycée (hormis à l'entrée en classe de 2nde et de 1ère année de CAP).

3. Méthodologie des évaluations standardisées

Les enquêtes mises en œuvre par la DEPP reposent sur les acquis de la psychométrie et s'inscrivent dans une démarche-qualité donnant lieu à une certification (AFNOR) en phase avec les standards internationaux. L'expérimentation est systématique avant toute mise en œuvre d'une évaluation standardisée. La conception d'une enquête repose sur la collaboration entre chercheurs et concepteurs de terrain formés (enseignants, conseillers pédagogiques, inspecteurs, etc.).

Les évaluations CP-CE1 sont construites dans un cadre de référence scientifique défini par le CSEN et explicité dans un document public (Évaluer pour mieux aider, CSEN). Des documents d'accompagnement des évaluations CP-CE1-6ème–2nde sont mis à disposition par la DGESCO sur le site Eduscol.

Toutes les des évaluations standardisées sont construites de façon à permettre des comparaisons diachroniques des acquis des élèves.

À l'exception des évaluations CP-CE1, l'ensemble des évaluations s'effectue désormais sur support numérique.

4. Qu'évaluent les évaluations standardisées ?

Les évaluations CP-CE1 et 6ème-2nde. En cohérence avec les priorités de la politique éducative actuelle, les évaluations sont principalement centrées sur l'acquisition des fondamentaux « lire-écrire-compter ». En ce sens, les évaluations sont à la fois un outil d'évaluation des acquis des élèves et un point d'appui de la mise en œuvre de la politique éducative. Au-delà des objectifs priorisés à une période donnée, elles permettent de s'assurer de la continuité des contenus d'évaluation en proposant une mesure des acquis des élèves et de leur évolution. Les liens avec les programmes d'enseignement sont explicites. Certains items (par exemple, la ligne numérique) de l'évaluation ont une visée prédictive de la réussite et du parcours d'apprentissage.

Le lien entre les programmes et les évaluations CEDRE est explicite.

Toutefois les évaluations standardisées portent quasi-exclusivement sur des compétences académiques. Les compétences transversales (socio-comportementales, socio-émotionnelles, fonctions exécutives, etc.) ne donnent pas, ou peu, lieu à évaluation. Elles sont pourtant largement prédictives de la réussite scolaire et de l'insertion socio-professionnelle des élèves. L'oral et les compétences physiques et artistiques constituent aussi un angle mort des évaluations standardisées. Toutefois, la DEPP commence à expérimenter des enquêtes sur le bien-être, la créativité et conçoit de nouveaux types d'enquêtes portant sur des observations de classe (interactions élèves-élèves et interactions élèves-professeurs).

Tableau 1 : Synthèse des évaluations standardisées réalisées en France

Niveau d'enseignement	Populations concernées – type d'évaluation	Objets d'évaluation - disciplines	Modalités de passation de l'évaluation – durée	Finalité(s)	Destinataires	Commanditaire-Conception	Périodicité
1 – ÉVALUATIONS NATIONALES EXHAUSTIVES							
Début CP	Ensemble des élèves – évaluation exhaustive	Programmes – Français Mathématiques	Évaluation écrite sous la forme d'un livret – correction par les professeurs (→ test unique) Français 3 séquences de 8 à 12 min Mathématiques : 2 séquences de 10 à 12 min	Diagnostic acquis élèves Appui pédagogique Questions prédictives Pilotage de proximité	Élèves Parents d'élèves Professeurs Inspecteurs de circonscription Remontées académique et nationale	MEN CSEN / DEPP / DGESCO / IGESR	Annuelle
Milieu CP	Ensemble des élèves – évaluation exhaustive	Programmes Français Mathématiques	Évaluation écrite sous la forme d'un livret – correction par les professeurs (→ test unique, différent chaque année) Français : 2 séquences de 8 à 10 min Mathématiques : 2 séquences de 11 min	Diagnostic acquis élèves Appui pédagogique Questions prédictives Pilotage de proximité	Élèves Parents d'élèves Professeurs Inspecteurs de circonscription Remontées académique et nationale	MEN CSEN / DEPP / DGESCO / IGESR	Annuelle
Début CE1	Ensemble des élèves – évaluation exhaustive	Programmes Français Mathématiques	Évaluation écrite sous la forme d'un livret – correction par les professeurs (→ test unique, différent chaque année) Durée : Français : 2 séq. Collectives de 12min et 2 séq. Individuelle d'1min Mathématiques : 2 séquences de 15 min	Diagnostic acquis élèves Appui pédagogique Questions prédictives Pilotage de proximité	Élèves Parents d'élèves Professeurs Inspecteurs de circonscription Remontées académique et nationale	MEN CSEN / DEPP / DGESCO / IGESR	Annuelle
Début 6ème	Ensemble des élèves – évaluation exhaustive	Programmes Français Mathématiques	Support numérique – 50 min	Test de positionnement individuel – diagnostic Appui pédagogique Pilotage	Élèves Parents d'élèves Professeurs Remontées académique et nationale	MEN / DEPP / DGESCO / IGESR	Annuelle Items libérés en partie**
Début 2ndeGT – 2nde PRO - CAP	Ensemble des élèves – évaluation exhaustive	Programmes Français Mathématiques	Support numérique – 50 min	Test de positionnement individuel – diagnostic Appui pédagogique Pilotage	Élèves Parents d'élèves Professeurs Remontées académique et nationale	MEN /DEPP / DGESCO / IGESR	Annuelle Items libérés en partie**
Journée Défense et Citoyenneté	Tous les jeunes avant 18 ans	Lecture Évaluation ponctuelle en numératie		Détection des jeunes en difficulté de lecture et de calcul	Indicateurs nationaux de suivi de performance en lecture + numératie	Ministère des armées / DEPP	Annuelle

* Les comparaisons dans le temps sont possibles pour certaines dimensions et pour toutes les dimensions entre 2019 et 2020 (stabilisation des épreuves), idem CE1 et mi-CP 2021 ce sera le cas également.

** Des items illustratifs sont libérés. En outre, à partir de cette année, deux épreuves (compréhension de l'écrit et résolution de problèmes) sont entièrement libérées (résultats de chaque élève à chaque item).

Niveau d'enseignement	Populations concernées Type d'évaluation	Objets d'évaluation - disciplines	Modalités de passation de l'évaluation	Finalité(s)	Destinataires	Commanditaire-Conception	Périodicité
2 - ENQUETES NATIONALES SUR ÉCHANTILLON							
CEDRE CM2	Échantillon représentatif (200 écoles – 6 000 élèves)	Programmes d'enseignement : Français / Mathématiques / HG + éducation civique / LVE / Sciences/ Attitudes à l'égard de la vie en société (2005)	Support numérique depuis 2019	Bilan des acquis des élèves Pilotage	Niveau national	MEN DEPP	5 ou 6 ans
CEDRE 3ème	Échantillon représentatif (310 collèges – 8 000 élèves)	Programmes d'enseignement : Français / Mathématiques / HG + éducation civique / LVE / Sciences/ Attitudes à l'égard de la vie en société (2005)	Support numérique depuis 2019	Bilan des acquis des élèves Pilotage	Niveau national	MEN DEPP	5 ou 6 ans
Socle commun CE2	Échantillon	Compétences et connaissances du socle commun Dimensions socio-cognitives Compétences du XXIème siècle (4C)			Niveau national (indicateurs LOLF – maîtrise du socle)	MEN	
Socle commun 6ème	Échantillon	Compétences et connaissances du socle commun Dimensions socio-cognitives Compétences du XXIème siècle (4C)	Support numérique	Bilan des acquis des élèves Pilotage	Niveau national (indicateurs LOLF – maîtrise du socle)	PARLEMENT/MEN DEPP	3 ans
Socle commun 3ème	Échantillon	Compétences et connaissances du socle commun Dimensions socio-cognitives Compétences du XXIème siècle (4C)			Niveau national (indicateurs LOLF – maîtrise du socle)	MEN DEPP	3 ans
Panels (6ème, 2007) (CP, 2011)	Échantillon	Parcours familial scolaire Et environnement		Suivi de cohortes	Niveau national	MEN DEPP	Panel
CM1 / 6ème	Échantillon – enquête ELAINE* 6 000 élèves de CM1 rentrés en 2019-2020 suivis jusqu'en 6e	Évaluation plan numérique	Test de compétences + entretien	Suivi de cohorte	Niveau national	MEN DEPP	
5ème / 3ème	Échantillon – enquête ELAINE 6 000 élèves de 5e rentrés en 2017-2018 suivis jusqu'en 3e	Évaluation plan numérique	Test de compétences + entretien	Suivi de cohorte	Niveau national	MEN DEPP	

Niveau d'enseignement	Populations concernées – type d'évaluation	Objets d'évaluation - disciplines	Modalités de passation de l'évaluation	Finalité(s)	Destinataires	Commanditaire /Conception	Périodicité
3 - ENQUETES INTERNATIONALES (échantillon)							
PISA (3ème / 2nde)	Élèves de 15 ans – (7 000 élèves)	Littératie / Compréhension de l'écrit / Culture mathématique / Culture scientifique	Numérique	Acquis des élèves Pilotage	Niveaux national et international	OCDE / DEPP	3 ans
PIRLS (CM1)	4 767 élèves (âge moyen des enquêtés : 10,2 ans ; France : 9,8 ans)	Compréhension de texte écrits	Numérique	Acquis des élèves Pilotage	Niveaux national et international	IAE / DEPP	5 ans
TIMSS (CM1)	6 000 élèves	Mathématiques et sciences	Tablette - papier	Acquis des élèves Pilotage	Niveaux national et international		1995 / 2015
TIMSS (4ème)	4 500 élèves	Mathématiques et sciences	Numérique	Acquis des élèves Pilotage	Niveaux national et international	IAE / DEPP	2019
TIMSS (TS)	4 000 élèves	Mathématiques et sciences	Papier	Acquis des élèves Pilotage	Niveaux national et international	IAE / DEPP	1995 / 2015
ICILS (4ème)	2 940 élèves (réponses complétées par des questionnaires aux professeurs)	Littératie numérique et pensée informatique		Acquis des élèves Pilotage	Niveaux national et international	IAE / DEPP	5 ans (Fr 2018)
ICCS (4ème)		Connaissances, attitudes et engagement civiques et citoyens des élèves		Acquis des élèves Pilotage	Niveaux national et international		Première participation de la France 2022

Tableau 2 : Parcours scolaire d'un élève et évaluations standardisées nationales et examens

Parcours scolaire		Types d'évaluations		
Cycles d'enseignement	**Années**	**Évaluations standardisées**		**Examens**
		Exhaustive	**Échantillon**	
Cycle 1	Petite section			
	Moyenne section			
	Grande section			
Cycle 2	CP	Début CP et mi-CP : Français et mathématiques – Évaluation annuelle		
	CE1	Début CE1 Français et mathématiques – Évaluation annuelle		
	CE2		Indicateur LOLF maîtrise du socle tous les 3 ans	
Cycle 3	CM1			
	CM2		CEDRE par discipline tous les 5 ou 6 ans	
	6ème	Test de positionnement : Français et mathématiques	Indicateur LOLF maîtrise du socle tous les 3 ans	
Cycle 4	5ème			
	4ème			
	3ème		CEDRE par discipline tous les 5 ou 6 ans Indicateur LOLF maîtrise du socle tous les 3 ans	Diplôme national du brevet / Attestation d'obtention du socle / Certificat de formation générale
Classe de détermination	2nde	Test de positionnement : Français et mathématiques		
Cycle terminal	1ère			Attestation de réussite intermédiaire (voie professionnelle) CAP (voie professionnelle)
	Terminale			Baccalauréat général, technologique ou professionnel

Thème 2 : Revue de littérature sur l'évaluation des acquis des élèves

Centrée sur les années 2000, la revue de littérature proposée ici repose sur un corpus de documents de différentes natures : textes officiels, rapports institutionnels, notes et études statistiques, articles et ouvrages scientifiques.

À partir de rapports des deux inspections générales du ministère de l'Éducation nationale (2005), de la Cour des comptes (2017) et de l'Assemblée nationale (2018), nous présentons tout d'abord un cadre général d'analyse de l'évaluation des acquis des élèves au sein du système éducatif (1), puis les résultats de travaux scientifiques et rapports relatifs aux évaluations standardisées (2), aux pratiques d'évaluation des professeurs (3) et aux examens (4).

1. Cadre d'analyse général de l'évaluation des acquis des élèves au sein du système éducatif

Le rapport conjoint IGEN-IGAENR intitulé « Les acquis des élèves, pierre de touche de la valeur de l'école ? »[2] conclut que « le système éducatif français s'intéresse, en 2005, d'une façon bien imparfaite et aléatoire à ce qui justifie aussi bien son existence que le fait que l'État lui consacre près du quart de ses ressources : les acquis des élèves qui le fréquentent ». Cette conclusion, étayée par une enquête des deux inspections générales auprès des directions centrales du ministère de l'éducation nationale et d'acteurs académiques, départementaux et d'établissement, repose sur les éléments suivants :

1. les outils de connaissance des acquis sont nombreux (résultats aux examens, données produites par la DEP[3] et l'OCDE, données académiques, etc.) mais sont soit trop rudimentaires (par exemple les taux de réussite aux examens ne renseignent que très peu sur les acquis effectifs des élèves), soit peu utilisés (par exemple les résultats aux évaluations nationales). Cette sous-utilisation des indicateurs d'évaluation des acquis des élèves est vérifiée à tous les niveaux, de l'établissement jusqu'aux directions centrales du ministère. Elle traduit une prise en compte insuffisante des acquis des élèves et la primauté d'une approche plus quantitative et gestionnaire que qualitative et pédagogique.
2. La cohérence entre les programmes d'enseignement, ce qui est effectivement enseigné dans la classe, et les acquis des élèves, est incertaine. Les modalités d'écriture des programmes sont cloisonnées entre les disciplines, et les programmes ne permettent pas toujours d'indiquer avec précision ce qui doit être enseigné. À l'exception de quelques études ponctuelles, les copies d'examen ne sont pas exploitées, alors même que leur analyse constituerait une ressource très riche pour évaluer les acquis des élèves.
3. L'évaluation des acquis des élèves dans « l'espace de la classe » reste, malgré des évolutions vers des formes plus formatives, très largement sommative et centrée sur l'octroi d'une note

[2] Rapport IGEN-IGAENR, *Les acquis des élèves, pierre de touche de la valeur de l'école ?*, n°2005-079, juillet 2005, p.59.
[3] Direction de l'évaluation et de la prospective (DEP) du ministère de l'éducation nationale qui deviendra en 2006 la Direction de l'évaluation, de la prospective et de la performance (DEPP).

souvent peu explicite et peu fidèle tant sa détermination est multifactorielle (classement des élèves entre eux, note pour encourager, etc.).

4. Enfin, le rapport insiste sur l'opacité et la grande hétérogénéité des informations données aux familles sur les acquis et résultats des élèves (par exemple hétérogénéité selon les ordres d'enseignement (école, collège, lycée) mais aussi au sein des ordres, hétérogénéité selon les disciplines et au sein des disciplines, etc.).

Sur la base de cette analyse, le rapport établit plusieurs recommandations visant à faire des acquis des élèves « un référentiel à partir duquel interroger l'ensemble du système éducatif ». Six axes sont alors proposés : établir « des programmes et des examens plus attentifs aux acquis des élèves », construire des « outils de lecture [des évaluations] des acquis adaptés aux besoins », « doter les élèves des tableaux de bord de leurs apprentissages », placer les acquis des élèves au cœur des pratiques des professeurs, développer de « nouvelles références d'inspection », évaluer les politiques à partir des acquis des élèves.

Depuis 2005, les lois d'orientation de 2005, 2013 et 2019 ont conduit à une évolution de nombreux éléments structurant de l'École : mise en place du Socle commun de connaissances, de compétences et de culture, création du Conseil supérieur des programmes, redéploiement des évaluations standardisées.

Cependant, deux rapports récents de la Cour des comptes (2017) et de l'Assemblée nationale (2018) reprennent assez largement les attendus du rapport de l'inspection générale de 2005. Ces deux rapports signalent en particulier :

- l'insuffisante prise en compte des acquis des élèves dont l'évaluation est « trop discontinue pour être utilisée » (*Rapport de la Cour des comptes*, p.37) et le fait que « l'éducation nationale n'a pas une vision claire sur ce que les élèves apprennent et sur la façon dont ils apprennent » (*Rapport de l'Assemblée nationale*, p.31),
- « une approche "traditionnelle" de l'évaluation fondée sur la notation chiffrée et centrée sur la performance et le classement des élèves » et des « évaluations en classe qui obéissent trop souvent à une logique de tri et de sanction » (*Rapport de l'Assemblée nationale*, p.61),
- des évaluations certificatives au « coût incertain » et qui « renseignent peu sur les acquis de connaissances et de compétences » (*Rapport de la Cour des comptes*, p.66),
- « une exploitation trop générale des résultats des évaluations des élèves, notamment de l'enquête PISA » (*Rapport de la Cour des comptes*, p.75-78).

Ces constats conduisent à des recommandations convergentes :

- « donner une plus grande cohérence à l'évaluation des acquis des élèves, celle-ci devrait être exhaustive et standardisée, se référer au socle commun et inciter les enseignants à adapter leurs pratiques pédagogiques » (*Rapport de l'Assemblée nationale*, p.37),
- « développer l'évaluation formative et positive en classe » (*Rapport de l'Assemblée nationale*, p.61),
- « mesurer systématiquement les acquis de la scolarité obligatoire » et « généraliser l'emploi des tests standardisés » (*Rapport de la Cour des comptes*, p.84-85),
- « approfondir les résultats des enquêtes internationales pour des enseignements concrets » (notamment en lien avec la critique du manque d'équité des résultats de l'École) (*Rapport de la Cour des comptes*, p.75-78).

Ces trois rapports, qui n'ont pas été sans effets sur certaines évolutions de la politique éducative et de l'organisation de l'école, conduisent à des conclusions assez similaires où il apparait que l'évaluation des acquis des élèves est mal assurée et insuffisamment prise en compte à tous les niveaux de l'École.

Faire de l'évaluation des acquis des élèves une donnée essentielle de l'évaluation de l'École invite à prendre en considération les attendus des travaux scientifiques sur les différentes modalités d'évaluation des élèves : les évaluations standardisées, les évaluations dans la classe et les examens.

2. Les évaluations standardisées

Mises en œuvre en France depuis une quarantaine d'années, les évaluations standardisées des acquis des élèves bénéficient, depuis le début des années 2000, d'un large écho qui s'étend bien au-delà des simples praticiens de l'École. Ces évaluations, qu'elles portent sur l'ensemble d'une population ou sur un échantillon, sont standardisées dans la mesure où leur conception, passation et correction sont strictement définies et réalisées dans des conditions identiques. Ainsi constituées, elles offrent une mesure objective des acquis des élèves et fournissent des éléments de comparaison fiables qui peuvent être exploités selon une double finalité :

- pédagogique : les résultats des évaluations donnent les moyens aux professeurs de mettre en œuvre des interventions pédagogiques adaptées aux besoins des élèves ;
- pilotage : les résultats des évaluations permettent d'éclairer le pilotage du système éducatif.

Classiquement, on distingue les évaluations diagnostiques, qui ont une vocation pédagogique, des évaluations bilans, qui sont avant tout des outils de pilotage.

Les sections suivantes proposent une revue de la littérature scientifique et de rapports institutionnels qui analysent : le recours croissant aux évaluations standardisées, la réception, les usages et effets de ce type d'évaluation et les évaluations standardisées internationales.

<u>Un recours croissant aux évaluations standardisées</u>

Le développement des évaluations standardisées des élèves s'observe en France (*Cf. infra*) comme au niveau international. Le recours croissant à ces formes d'évaluation depuis les années 1990 est présenté dans la littérature scientifique (Maroy, 2008 ; Dupriez et Malet (dir.), 2013) comme un des effets de l'application dans le champ des politiques d'éducation de nouvelles conceptions et modalités de mise en œuvre des politiques publiques (*new public management*) qui se sont traduites par :

- un renforcement de l'autonomie et de la responsabilité données aux acteurs (mouvements de déconcentration et de décentralisation),
- une attention plus forte à l'efficacité et l'efficience des mesures adoptées, notamment au travers la mise en œuvre de politiques d'évaluation,
- un développement de la redevabilité (*accountability*) de l'École envers ses usagers et le grand public.

Ces politiques d'évaluation standardisées sont désormais largement répandues dans la plupart des pays développés. En 2009, la synthèse établie par le réseau Eurydice (Eurydice, 2009) indique qu'au sein de l'Union européenne seulement cinq pays ou régions n'ont pas d'évaluation nationale

(Communauté germanophone de Belgique, Grèce, Liechtenstein, Pays de Galles et République tchèque).

Les politiques d'évaluations standardisées des élèves ne sont toutefois pas identiques dans tous les pays. On distingue deux grands modèles (Dupriez et Mons, 2010) :

- le modèle anglo-saxon (« *accountabiliy* dure ») qui repose sur des tests à forts enjeux : leurs résultats sont publics et peuvent avoir des conséquences importantes sur les établissements scolaires (fermeture d'école, octroi de subvention, etc.) et le parcours des élèves (redoublement) ;
- le modèle de l'Europe continentale (« *accountabiliy* réflexive ») pour lequel l'utilisation des résultats n'a pas d'effets directs sur les professeurs, les élèves ou les établissements scolaires. L'objectif est ici que la diffusion des résultats agisse comme un feedback ou un effet miroir (Thélot, 2002) en confrontant les acteurs aux effets de leurs actions.

Entre ces deux modèles, on retrouve selon les pays des variantes et des configurations plus ou moins hybrides, par exemple en ce qui concerne le degré de diffusion des résultats des établissements ou le lien entre la formation continue des enseignants et les évaluations standardisées (Dupriez et Mons, 2010).

De façon générale, le recours aux évaluations externes standardisées se développe. C'est le cas de la Suisse où des évaluations visant à vérifier le degré d'atteinte des compétences fondamentales (tests COFO) ont été mises en œuvre depuis 2016.

Toutefois, certains pays (Corée du Sud, Angleterre) ont récemment allégé leurs dispositifs d'évaluation des élèves. En Corée du Sud, il s'agit de faire évoluer un système très axé sur les tests en développant un curriculum recentré sur la promotion de la créativité et en donnant la priorité à l'expression écrite plutôt qu'aux questionnaires à choix multiples. En Angleterre, la conception de certaines évaluations n'est plus centralisée mais revient aux professeurs tandis que d'autres évaluations sont devenues facultatives.

Réceptions, usages et effets des évaluations standardisées des acquis des élèves

Une réception des évaluations parfois controversée

De nombreux travaux se sont intéressés à la réception et aux usages des évaluations standardisées par les enseignants, notamment pour tenter de comprendre les réticences voire les résistances dont elles font (ou ont fait) l'objet dans différents pays.

En procédant à une analyse sociologique et historique, Xavier Pons (Pons, 2012) considère que l'histoire des évaluations standardisées se caractérise en France depuis 40 ans par un double mouvement de complexification et de politisation.

Elles se sont complexifiées en raison de l'utilisation d'une méthodologie de plus en plus sophistiquée et d'une « architecture institutionnelle d'ensemble qui, loin d'avoir connu un mouvement de rationalisation technique s'est plutôt traduite par un empilement de dispositifs aux finalités multiples » (Pons, 2012, p.17). On assiste effectivement depuis les années 1990 à une succession d'évaluations

standardisées parfois non pérennes[4] et dont les modalités et finalités sont parfois contestées. En conjuguant les caractéristiques d'une évaluation bilan à vocation de pilotage et d'une évaluation diagnostique à vocation pédagogique, les évaluations mises en œuvre de 2008 à 2012 auprès de l'ensemble des élèves de CE1 et CM2 ont suscité de vives polémiques et de l'incompréhension entre les professeurs et le ministère qui était accusé de vouloir utiliser ces évaluations pour évaluer les personnels et classer les écoles. L'approche de l'élection présidentielle de 2012 a contribué à fortement politiser ce conflit que Xavier Pons qualifie de « corporatiste classique » (*Ibid*., p.16).

L'analyse de la réception des évaluations CE1-CM2 développée par les sociologues Yves Dutercq et Xavière Lanéelle suggère que les évaluations externes des élèves peuvent se comprendre pour les professeurs « comme une épreuve » au sens sociologique, c'est-à-dire « comme un moment où les compétences et la valeur (la grandeur) d'un individu sont publiquement mises en question, en particulier quand cette valeur s'est construite en référence à un principe ou à un ensemble de principes (un monde) qui est dénoncé en cette occasion au nom d'un autre principe (d'un autre monde) » (Dutercq et Lanéelle, 2013, p.44- 45).

Selon cette perspective, les tensions sont donc l'expression d'un conflit portant sur des valeurs et des principes de justice qui engagent des représentations opposées de l'école et du métier de professeurs.

Au-delà de la seule réception des évaluations standardisées des acquis des élèves, Lise Demailly conçoit la résistance des enseignants aux évaluations comme une « énigme sociologique » (Demailly, 2003, p.117) soulevant plusieurs paradoxes. En effet, les enseignants évaluent eux-mêmes constamment leurs élèves. De plus, l'évaluation des élèves est aussi un moyen de reconnaissance du travail de formation que les enseignants effectuent.

Pour lever cette énigme, qui « n'est ni totale ni constante, mais semble plutôt varier selon les contextes et les conjonctures » (Ibid.), Lise Demailly avance trois arguments principaux. Premièrement, la résistance à l'évaluation s'explique par la défense d'intérêts individuels et corporatistes. Deuxièmement, elle est l'expression d'un conflit de valeurs (les évaluations désacralisent l'école et renvoient à un management d'entreprise) et d'un sentiment d'injustice (les évaluations servent à faire des comparaisons qui sont injustes car les conditions de travail et les contextes scolaires sont tous différents). Enfin, la résistance aux évaluations standardisées réside dans une méfiance envers toutes formes d'opérations statistiques considérées comme facilement manipulables.

En synthétisant différents travaux sociologiques et institutionnels des années 2000, Lise Demailly présente des éléments de convergence qui peuvent permettre de dépasser certaines oppositions à l'évaluation. Il faut tout d'abord considérer que « l'évaluation n'est pas qu'une pratique technique, c'est une pratique stratégique et politique » (*Idib*., p.116) puis que le développement « d'une "culture de l'évaluation" ne peut consister en la simple amélioration de la rigueur et de la finesse intrinsèque des évaluations » (*Idib*., p.116). En conséquence, « les évaluations doivent être participatives si elles veulent mobiliser les acteurs et faire en sorte que la décision publique ou professionnelle soit effectivement orientée par l'évaluation » (*Idib*., p.116).

[4] En 2001, l'évaluation diagnostique à l'entrée en 2nde ; en 2002 une évaluation de ce type est instaurée en 5ème ; en 2004-2005, une évaluation diagnostique en CE1 est expérimentée, reconduite en 2005-2006, généralisée en 2006-2008 avant d'être supprimée en 2008.

En France, après plusieurs années de polémiques vives, les évaluations standardisées sont désormais bien acceptées par les enseignants. En 2012, un rapport commun de l'IGEN et de la DEPP[5]5 indique qu'après « avoir été contestées les évaluations semblent mieux acceptées sur le terrain que par les représentants des enseignants que ce soit au CE1 ou au CM2. » (p.2).

En 2019, un rapport de l'IGÉSR[6] confirme ce constat dans le premier degré où « le regard des équipes est généralement positif sur les évaluations. Elles sont globalement appréciées et leur utilité est reconnue tout particulièrement quand les enseignants ont pu disposer de repères par rapport à la circonscription ou au département leur permettant de cerner, par comparaison, les points de force et de faiblesses de leurs élèves. » (p.6). En revanche dans le second degré, « les enseignants accueillent les évaluations sans opposition, mais l'intérêt qu'ils y portent reste limité »[7] (p.11), le rapport conclut que « les enseignants de collège espèrent que les évaluations leur seront utiles, alors que ceux de lycées sont, dès les passations, convaincus de l'inutilité du dispositif. » (*Idib*., p.12).

Au-delà de la seule réception par les professeurs des évaluations standardisées, qu'en est-il de leurs usages et quels sont leurs impacts sur les pratiques pédagogiques et la réussite des élèves ?

Usages et impacts sur les pratiques pédagogiques et les résultats de élèves

Dans un article publié en 2009, Nathalie Mons (Mons, 2009) établit une large revue de littérature sur « les effets théoriques et réels des politiques d'évaluation standardisée ». Sa principale conclusion est qu'il existe un écart entre les effets théoriques de ces politiques et leurs effets observés.

Quels sont les effets attendus ? Sur le plan théorique, les effets attendus des évaluations standardisées (*testing*) diffèrent selon la façon dont les résultats aux tests sont utilisés dans la régulation des systèmes éducatifs.

Dans le modèle anglo-saxon, les effets attendus sont :

- pour les élèves, de meilleurs apprentissages liés à un travail plus important et un surcroît de motivation (peur du redoublement, volonté d'accéder à la classe supérieure ou à une certification),
- pour les professeurs, une responsabilisation, un professionnalisme et une motivation renforcés par la publication des résultats de leur école,
- pour les responsables administratifs : une gestion plus efficace en raison d'une meilleure connaissance des résultats de chaque école,
- une réduction des inégalités scolaires liée à l'application de standards communs à tous les élèves.

[5] Rapport IGEN-DEPP, *Suivi qualitatif des conditions de passation et de correction des évaluations nationales exhaustives de CE1 et de CM2 en 2012*, n°2012-98, juillet 2012.
[6] Rapport de l'IGÉSR, *L'organisation et la passation des évaluations nationales dans le premier degré*, n°2019-096, octobre 2019.

[7] Rapport de l'IGÉSR, *Les évaluations nationales du second degré*, n°2020-074, juin 2020.

Dans le modèle de l'Europe continentale, les effets attendus sont en lien avec l'introduction de « standards de résultat » (c'est-à-dire des objectifs d'enseignement à atteindre) doit permettre :

- de clarifier les attendus de l'enseignement pour l'ensemble des acteurs (professeurs, familles, élèves),
- de renforcer l'efficacité du travail pédagogique des professeurs en leur donnant les moyens de s'interroger sur les effets de leurs actions. Cette analyse correspond à l'effet-miroir théorisé en France par Claude Thélot (Thélot, 2002).

Quels effets observés sur les résultats des élèves ? L'analyse des réformes menées au Texas[8] et dans le district de Chicago[9] montre des résultats sur le niveau des élèves très ambivalents, voire contradictoires. Dans les deux cas, les premiers résultats montrent une élévation importante des résultats des élèves. Ces données sont toutefois contestées par d'autres études qui révèlent l'inconsistance de ces résultats dans le temps, leur variabilité selon les disciplines et la relative faiblesse des progrès lorsqu'ils sont comparés à ceux observés dans des tests nationaux.

Plusieurs études sur lesquelles s'appuient la synthèse de Nathalie Mons (Mons, 2009) montrent que l'élévation des résultats s'explique par la mise en œuvre de stratégies d'entraînement aux tests (*teaching to the test*), d'un renforcement de la sélection des élèves (exclusion des plus faibles) et d'une polarisation des curricula sur lesquels il est plus aisé de faire progresser les élèves (par exemple en mathématiques : le calcul mental est privilégié à la résolution de problèmes).

D'autres études s'interrogeant sur la relation entre l'intensité de l'*accountability* et les performances des élèves concluent de la même façon à l'incertitude des résultats : des mesures d'*accountability* plus prégnantes ne s'accompagnent pas systématiquement d'une élévation des résultats des élèves. Il ne semble donc pas exister de relation stable et significative entre les dispositifs d'*accountability* et les performances des pays.

Toutefois, plusieurs études montrent que « l'organisation d'examens nationaux externes dans l'enseignement secondaire semble être en lien avec des inégalités plus faibles » (Mons, 2009, p.117), notamment parce qu'ils permettent d'homogénéiser les « exigences académiques » (*Ibid*.) et de limiter « les dérives curriculaires dans les établissements qui accueillent les élèves des milieux défavorisés. » (*Ibid*.).

La méta-analyse de 14 études internationales conduite par le chercheur américain Jaekyung Lee confirme l'ambivalence des effets empiriques des évaluations standardisées : les politiques de test peuvent améliorer les résultats des élèves mais il subsiste de fortes disparités selon les études (quelles que soient les disciplines considérées). Enfin, ces politiques n'ont pas d'effets univoques sur la réduction des inégalités scolaires.

[8] Au début des années 1990, l'État du Texas procède à une politique de tests d'évaluation des acquis à différents niveaux de la scolarité. En fonction des performances des élèves aux tests les écoles sont classées et leur classement donne lieu à un système de récompenses (octroi de financement) et de sanctions (risque de fermeture). Le parcours des élèves dépend de leurs résultats aux tests (redoublement, obtention d'un diplôme).
[9] En 1995, le district de Chicago conditionne le passage à plusieurs niveaux d'enseignement aux résultats des élèves à des tests standardisés.

Quel bilan de la politique américaine No child left behind ?

La synthèse des études sur les effets du programme *No child left behind* (NCLF) proposée par Denis Meuret (Meuret, 2015) offre une analyse nuancée des effets des politiques d'évaluation standardisées à forts enjeux : si le programme est relativement neutre sur les performances globales des élèves, l'équité scolaire a progressé (le test PISA montre une élévation des résultats des élèves les plus faibles et une baisse des résultats scolaires des meilleurs élèves).

L'étude nuance également quatre critiques habituelles :

- creusement des inégalités : la relation est positive (au pire nulle) sur la baisse des inégalités,
- enseignement au rabais (*teaching for the test*) : l'effort a bien été mis sur les matières évaluées (c'était l'objectif de la loi), mais les autres disciplines, non évaluées, n'en ont pas pâti,
- attention plus soutenue accordée aux élèves les plus proches du niveau de réussite : cette crainte s'observe effectivement pour les tests à standards élevés ; « si les standards sont moins élevés alors les inégalités se réduisent » (Meuret, 2015, p.4),
- baisse du moral des enseignants : les études montrent en réalité que « l'effet net de NCLB a été positif sur le sentiment des enseignants de contrôler leur classe, d'être soutenus par l'administration et nul ou non significatif sur leur charge de travail, sur leur satisfaction professionnelle comme sur leur envie de rester dans la profession. » (*Ibid*., p.5).

Quels effets observés sur les pratiques pédagogiques ? En synthétisant différentes enquêtes, Nathalie Mons indique que « ces réformes permettent […] de donner des guides clairs pour mettre en œuvre les curricula, d'empêcher l'apparition de fortes inégalités dans le développement des syllabi locaux, de mettre l'accent sur les résultats réels des élèves, en particulier pour ceux issus des milieux défavorisés, et de favoriser un travail en équipe autour de l'analyse des résultats des évaluations. » (Mons, 2009, p.121).

Si les effets globaux semblent positifs, des effets pervers peuvent toutefois exister sous la forme de phénomènes de *teaching to the test*, d'un rétrécissement des contenus d'enseignement, d'une focalisation sur certaines compétences utiles aux épreuves (par exemple, la mémorisation) au détriment d'autres jugées moins indispensables (créativité, autonomie), d'une centration sur les objectifs d'apprentissages cognitifs aux dépens d'autres missions de l'École (autonomie, citoyenneté, etc.). Ces effets sont d'autant plus fréquents que les épreuves sont à fort enjeu.

L'enquête statistique conduite par Esteban Rozenwajn et Xavier Dumay (Rozenwajn et Dumay, 2016) permet d'actualiser la compréhension des effets des évaluations standardisées sur les pratiques d'enseignement[10], en précisant les changements survenus dans les pratiques pédagogiques et en cherchant à identifier les processus cognitifs à l'œuvre dans le traitement des résultats aux évaluations externes.

[10] Il s'agit d'une enquête par questionnaire auprès d'enseignants belges du secondaire inférieur confrontés à l'épreuve certificative de la fin du 1er degré (14-15 ans). L'épreuve a un enjeu pour les élèves puisqu'elle conditionne leur accès aux études supérieures. En revanche, elle n'a pas d'effets formels sur les enseignants et les établissements.

Les données de l'enquête confirment le faible impact des résultats des évaluations sur les pratiques pédagogiques : les changements observés sont bien davantage « instrumentaux et curriculaires que centrés sur les méthodes pédagogiques proprement dites. » (Rozenwajn et Dumay, 2016, p.72). En outre, parmi les processus cognitifs testés, c'est l'attitude des enseignants (c'est-à-dire leur inclinaison initiale) envers les évaluations externes et l'influence sociale des collègues (c'est-à-dire l'estimation de l'importance que leurs collègues accordent aux évaluations) qui jouent un rôle central dans l'utilisation des tests à des fins de changement pédagogique. En d'autres termes, « l'effet miroir » escompté semble peu opérant.

Quelles conditions d'appropriation des évaluations externes par les enseignants ? À partir de l'analyse de la façon dont les enseignants utilisent les résultats de leurs élèves aux évaluations de CE2 et de 6ème (en vigueur de 1989-1990 à 2007-2008), Jean-Louis Derouet et Romuald Normand concluent que : « les difficultés de diffusion et de traduction des évaluations nationales s'expliquent par le fait qu'elles mettent en tension des mondes d'acteurs différents, avec leur propre rationalité et leur propre légitimité : celui des responsables institutionnels soucieux de faire des évaluations nationales un instrument de prise de décision et de pilotage de leur action, celui des évaluateurs attachés à la validité et à la pertinence scientifique des outils mis en œuvre, celui des professionnels de la formation intéressés dans la diffusion et le partage des acquis de la didactique et de la recherche en éducation, celui des chefs d'établissements préoccupés de satisfaire les exigences de leur hiérarchie mais devant composer avec le contexte local de leur collège, celui des enseignants pour qui ces évaluations nationales sont étrangères à leur culture professionnelle et au sens qu'ils donnent à leur travail pédagogique avec les élèves. » (Derouet et Normand, 2003, p.188).

Dès lors, l'enjeu est d'être en mesure de créer « un espace d'intéressement entre tous les acteurs » (*Idid*., p.189) afin de réduire « les tensions entre une logique de pilotage où les évaluations nationales servent d'indicateurs pour la prise de décision et une logique d'accompagnement à la mise en œuvre de dispositifs d'innovation et de remédiation pédagogique » (*Ibid.*), notamment en clarifiant les enjeux des dispositifs d'évaluation mis en place, en multipliant les occasions d'une appropriation au plus près la classe, des pratiques pédagogiques, des résultats et outils de remédiation.

Il y a très peu d'études sur la façon dont les élèves réceptionnent et « vivent » ces évaluations. Les études les plus critiques insistent sur le stress et les effets négatifs des tests à fort enjeu sur le parcours scolaire des élèves.

Une autre approche consiste à s'interroger sur la motivation des élèves à répondre à des évaluations à faible enjeu. En effet, une faible motivation à répondre peut nuire à la validité des résultats et à leur comparabilité. Une étude de la DEPP en 2015 (Keskpaik et Rocher, 2015) conclut à l'importance des conditions de passation et de l'explicitation des enjeux auprès des élèves. Les auteurs observent également que la motivation des élèves face au test varie selon leurs caractéristiques : les garçons et les redoublants sont relativement moins impliqués et la motivation est plus faible en éducation prioritaire. Toutefois, la relation entre la motivation à répondre et la performance est limitée. De même la motivation varie peu d'un établissement à l'autre.

Apports et limites des évaluations internationales

La première enquête internationale d'évaluation des acquis des élèves, menée par l'Institut international de l'éducation, date de 1961. Mais c'est essentiellement depuis les années 2000 que la comparaison internationale s'est développée à travers la création des enquêtes de l'IEA[11] (TIMSS en 1995, PIRLS en 2001) et de l'OCDE (PISA en 2000).

Les principales vertus de ces enquêtes sont d'offrir des éléments de comparaison entre les systèmes éducatifs nationaux et d'objectivation de leurs performances en termes d'acquis des élèves. Elles peuvent également servir de cadre de référence pour la mise en œuvre de réformes. C'est notamment le cas de l'Allemagne où la déception provoquée par les résultats de PISA 2000 a entraîné de profondes réformes du système éducatif.

En France, il n'y a pas eu à proprement parler de « choc PISA », même si de nombreuses productions scientifiques ont permis de révéler les caractéristiques, forces et difficultés du système éducatif que les enquêtes PISA permettent de révéler (Baudelot et Establet, 2009). De façon très synthétique :

- les résultats des élèves âgés de 15 ans qui se situent globalement dans la moyenne des pays de l'OCDE ;
- l'origine sociale et les résultats des élèves sont très corrélés ;
- il existe un déficit dans la maîtrise de certaines compétences socio-comportementales (Algan, Huillery, Prost, 2018).

Les enquêtes internationales présentent des vertus heuristiques largement reconnues : « l'enquête PISA est d'une très grande qualité et offre peu de prises à une critique purement méthodologique » (Grenet, 2008, p.8). Cependant de nombreuses analyses, qui portent le plus souvent sur l'enquête PISA, insistent sur les débats méthodologiques (Revue française de pédagogie, n°157, 2006/4) et les tentations « surinterprétatives » (Duru-Bellat, 2012) que cette enquête suscite.

Ces études montrent que le cadre théorique retenu (types de questions, définition des compétences évaluées, supports documentaires utilisés, etc.) peut induire des « biais culturels » difficiles à neutraliser.

La principale source de biais selon Julien Grenet réside dans la composition des échantillons d'élèves. L'enquête PISA cible la population âgée de 15 ans scolarisée, or les taux de scolarisation à 15 ans diffèrent fortement selon les pays, ce qui entraîne des biais statistiques importants et limite la portée des comparaisons. Celles-ci doivent en premier lieu être effectuées entre pays ayant des taux de scolarisation à 15 ans comparables.

En outre, toute comparaison reposant sur des échantillons représentatifs d'une population induit des marges d'erreur dans l'interprétation des résultats. Dans le cadre de l'enquête PISA, on estime ainsi que les résultats d'un pays sont affectés d'une marge d'erreur de plus ou moins 5 points (soit une plage de 10 points). Dès lors, une attention trop étroite au classement d'un pays peut donner une image trompeuse de ses résultats.

[11] International Association for the Evaluation of Educational Achievement.

L'un des objectifs de l'enquête PISA est aussi de promouvoir des bonnes pratiques et d'aider les pays à améliorer les résultats de leurs élèves. Les enquêtes PISA peuvent donc s'inscrire dans une certaine normativité qui, lorsqu'elle s'applique trop directement, fait fi des singularités institutionnelles de chaque pays.

Dans cette perspective, Marie Duru-Bellat invite à se méfier des risques de surinterprétation des résultats, en particulier quand les corrélations se transforment trop rapidement en causalité (Duru-Bellat, 2012). Dans le domaine de l'éducation, il est toujours très risqué d'imputer à un seul facteur tel ou tel résultat.

Certains auteurs appellent à établir clairement une différence entre « ce que disent les enquêtes PISA et ce qu'on en dit » (Felouzis et Charmillot, 2012) qu'il s'agisse des recommandations proposées par l'OCDE ou des commentaires et angles de vues adoptées par les différents acteurs politiques, syndicaux et médiatiques.

Cherchant à comprendre pourquoi la France n'a pas connu, à la différence d'autres pays, de « choc PISA » malgré des résultats que d'aucuns jugent décevants, certains auteurs montrent que la réception des enquêtes PISA est liée à des configurations politiques, institutionnelles, professionnelles qui conditionnent sa mise à l'agenda et son impact sur la politique publique (Mons, Pons, 2013).

3. L'évaluation dans la classe et les examens

Bien que l'évaluation des acquis des élèves soit au cœur de l'enseignement, les pratiques d'évaluation des enseignants ont moins fait l'objet de publications scientifiques ces dernières années que les évaluations standardisées et d'autres formes d'évaluation de l'École (CNESCO 2014).

Plusieurs études et rapports institutionnels permettent toutefois de dresser un portrait des évaluations dans la classe telles qu'elles sont pratiquées en France.

Les pratiques d'évaluation à l'école primaire

Les pratiques d'évaluation à l'école primaire ont fortement évolué au cours des années 2000. Désormais l'évaluation chiffrée n'intervient qu'à partir du cycle 3. Ainsi « la notation chiffrée apparaît, à l'école primaire, comme un épiphénomène dans ce qui différencie le système français par rapport à d'autres systèmes éducatifs » (IGEN, 2013, p.4).

Cette évolution est analysée comme la lente diffusion de certaines préconisations pédagogiques, la mise en place du socle commun et de son évaluation via le Livret Personnel de Compétence (LPC) puis, depuis 2017, le Livret Scolaire Unique (LSU), qui organise une évaluation des compétences en fonction d'objectifs à atteindre (non atteints, partiellement atteints, atteints, dépassés[12]).

[12] Le rapport signale que cet item « objectifs dépassés » suscite beaucoup d'incompréhension tant de la part des parents d'élèves que des professeurs. Au collège, le positionnement s'effectue à partir de degrés de maîtrise (insuffisante, fragile, satisfaisante, très bonne). « Ce changement de terminologie dont on comprend mal les raisons a déstabilisé les familles. » (IGEN, 2017, p.10).

L'évaluation des acquis des élèves reste toutefois différente selon les cycles. Au cycle 1, conformément aux recommandations du programme[13], l'évaluation se fonde sur l'observation et vise avant tout les progrès des élèves. La notation n'existe pas en petite et moyenne sections et est très exceptionnelle en grande section. Toutefois, les modalités utilisées pour rendre compte des acquis et progrès peuvent être très variées (code couleurs, émoticônes, icônes, etc.).

Encore peu présente au cycle 2, la notation (en plus des appréciations littérales) devient plus fréquente au cycle 3 pour les évaluations liées au suivi quotidien (en particulier en CM1 et CM2). En revanche, les bilans périodiques renseignés par l'intermédiaire du LSU ne comportent pas de notes.

La mission de l'IGEN de 2013 signale que l'évaluation à l'école primaire développe des « dispositifs multiples et concomitants, manquant souvent d'unité » (IGEN, 2013, p.7). Ainsi, pour une même école, les démarches d'évaluation sont parfois concertées, par exemple pour adopter des codes d'évaluation identiques (lettres, couleurs, etc.), mais il existe également des situations où des dispositifs très différents coexistent.

Ces situations sont l'expression d'un manque de continuité et de cohérence entre les professeurs. Elles sont source de complexité et d'incompréhension pour les élèves et leurs parents. Au-delà de l'intérêt d'adopter un dispositif d'évaluation cohérent entre les professeurs d'une même école (ou *a minima* d'un même niveau), le rapport souligne que « les enseignants se concertent sur l'outil pour rendre compte, mais, rarement, sur les critères ou les seuils de performance attendus, ce qui entraîne une objectivité avant tout formelle et de façade. » (*Ibid*., p.14).

Les pratiques d'évaluation dans le second degré

Les enseignants des collèges et lycées consacrent en moyenne 5h48 heures par semaine à la correction et à l'évaluation des élèves, soit près 26% de leur temps de travail en dehors des cours, des différences importantes existant selon les niveaux et disciplines enseignées (DEPP, 2013).

L'étude de la DEPP sur les pratiques d'évaluation des enseignants au collège (DEPP, 2004) indique que l'évaluation est essentiellement sommative même si des évolutions vers des évaluations plus formatives et diagnostiques sont perceptibles. Les objets des évaluations concernent avant tout les compétences disciplinaires (citées dans 97% des réponses des enseignants enquêtés) et les connaissances (84%), les compétences transversales sont moins fréquemment évaluées (31%).

Plus précisément, l'enquête de la DEPP établit une typologie des pratiques d'évaluation des enseignants en distinguant trois groupes :

- Le groupe 1 rassemble 50% des professeurs enquêtés, qui enseignent principalement des disciplines scientifiques. Ils sont globalement « peu soucieux de différenciation ou de remédiation, ces enseignants décident seuls des modalités de l'évaluation, à partir des manuels. Leurs pratiques évaluatives, centrées sur les contenus disciplinaires, s'appuient sur

[13] Extrait du programme du cycle 1 : « Une école qui pratique une évaluation positive. L'évaluation constitue un outil de régulation dans l'activité professionnelle des enseignants ; elle n'est pas un instrument de prédiction ni de sélection. Elle repose sur une observation attentive et une interprétation de ce que chaque enfant dit ou fait. Chaque enseignant s'attache à mettre en valeur, au-delà du résultat obtenu, le cheminement de l'enfant et les progrès qu'il fait par rapport à lui-même. »

des écrits de restitution, dans une démarche d'évaluation plutôt sommative, située en fin de séquence d'apprentissage » (DEPP, 2004, p.4) ;

- Le groupe 2 représente 38% de l'échantillon et regroupe principalement des professeurs de lettres, langues et histoire-géographie. Leurs réponses « révèlent un souci de diversification des ressources, des objectifs, des modalités de l'évaluation, une approche plus collective à partir des différents critères : appel à l'autonomie et à la créativité des élèves dans des tâches qui sollicitent des compétences transversales prises en compte dans les barèmes de notation, élaboration des évaluations au fur et à mesure des séquences en fonction du niveau d'acquisition réel des élèves ; importance donnée à la correction, au repérage des erreurs, à l'implication des élèves dans cette phase ; utilisation des résultats des évaluations pour individualiser les apprentissages, proposition de remédiations, réorganisation des contenus enseignés, etc. » (*Ibid*., p.5) ;
- Le groupe 3 représente 12% de l'échantillon et comprend en grande majorité des professeurs d'EPS et quelques professeurs d'arts plastiques. « Systématiquement placées en début de séquence d'apprentissage, leurs évaluations sont élaborées collectivement et portent surtout sur des savoir-être. Elles sont de nature diagnostique et formative puisqu'elles sont souvent utilisées pour constituer des groupes de besoin et individualiser les apprentissages. » (*Ibid*., p.5).

Globalement en France, les pratiques d'évaluation restent plutôt individuelles : 95% des enseignants indiquent fixer eux-mêmes le calendrier et le rythme de leur progression. L'enquête TALIS 2013 nuance ce constat, mais la collaboration entre les enseignants demeure nettement plus faible en France que dans les autres pays de l'OCDE (CNESCO 2014) : 20% des enseignants du premier cycle secondaire déclarent ne jamais collaborer avec d'autres enseignants pour appliquer des barèmes communs et évaluer les progrès des élèves contre seulement 8% en moyenne dans les pays ayant participé à l'enquête TALIS.

La conception des évaluations est aussi plus individuelle en France que dans les autres pays de l'enquête TALIS (86% des enseignants du premier cycle secondaire déclarent réaliser leurs propres contrôles écrits « souvent » ou à « chaque séance » contre 68% en moyenne dans les pays étudiés). La France se distingue également par la faible diffusion de l'auto-évaluation, relativement aux autres pays de l'enquête TALIS.

En revanche, l'usage de l'oral est un peu plus développé en France que dans la moyenne des autres pays de l'enquête TALIS et l'observation des élèves est presque aussi fréquente en France que dans les pays de l'échantillon TALIS 2003.

La notation reste très présente au collège. Elle est la plupart du temps assortie d'un commentaire : 75% des enseignants déclarent ajouter régulièrement un commentaire écrit à la note chiffrée. Ils sont en moyenne 54% à le faire dans les pays de l'échantillon TALIS.

Cependant, les pratiques d'évaluation au collège tendent à évoluer et à se diversifier au travers (1) de l'introduction du socle commun (entrée en vigueur progressive au collège à partir de 2009) et (2) du développement de nombreuses expérimentations, par exemple les dispositifs de classes sans note.

1. La validation du socle commun de compétences, de connaissances et de culture est devenue un élément essentiel de l'évaluation au collège et de l'obtention du DNB. Les différents rapports de l'inspection générale (IGAENR-IGEN, 2012 ; IGEN, 2017) dressent un bilan contrasté de la mise en œuvre du LPC puis du LSU[14].

   L'introduction de ces livrets a effectivement permis d'engager une réflexion sur les pratiques d'évaluation et leur nécessaire harmonisation, particulièrement au sein d'un même établissement. Mais, au-delà des difficultés techniques de déploiement de l'outil numérique du LSU et de son inter-opérationnalité avec les logiciels d'édition des bulletins scolaires, l'inspection générale souligne d'une part les difficultés des professeurs à articuler les résultats de leurs élèves dans des évaluations disciplinaires avec l'attribution d'un niveau de maîtrise dans un domaine du socle[15], d'autre part la grande diversité selon les collèges des modalités d'attribution des niveaux de maîtrise des domaines du socle et de la répartition de ces domaines entre les différentes disciplines.

   Cette variabilité entre les établissements n'est pas sans effet sur « les niveaux de maîtrise attribués des domaines du socle, ce qui soulève la question de l'équité d'une évaluation qui impacte à la fois la note obtenue à un examen national (le DNB, séries générales et professionnelles) et l'affectation post-troisième. » (IGEN, 2017, p.12).

2. Les expérimentations du type « classes sans note » se sont multipliées depuis les années 2010. Elles constituent un élément important de l'évaluation des pratiques évaluatives au collège. Il est difficile d'en tirer un bilan univoque. L'inspection générale (IGEN, 2013) constate qu'une des vertus de cette démarche réside dans l'instauration d'une dynamique de projet pédagogique et didactique qui facilite le travail en commun des professeurs sur la notion d'évaluation.

   Du côté des élèves, les effets positifs sont une participation plus importante, notamment des élèves les plus en difficulté, une meilleure estime de soi et une motivation renforcée (voir également, Merle 2012 et 2018). Toutefois selon la mission de l'IGEN, il est difficile « de savoir si ces effets relèvent du dispositif en lui-même ou d'un changement dans la démarche pédagogique » (IGEN, 2013, p.19) ou bien d'un « investissement renforcé d'équipes particulièrement motivées et mobilisées » (*Ibid*.)

   Enfin, la mission relève plusieurs difficultés qui entravent la mise en place de ces dispositifs et leur pérennité : « les résistances liées aux représentations (à commencer par celles des parents) qui prévalent dans la société, les limites liées à l'environnement professionnel et au statut des professeurs, enfin celles liées au système éducatif lui-même, qui ne parvient pas à accompagner ou à encadrer suffisamment les innovations observées. »

   Une étude économétrique (Algan, Constantin, Delpeuch, Huillery, Prost, 2018) testant l'impact de l'introduction d'une évaluation par compétence (classe sans note) auprès de 89 établissements conclut à l'absence « d'effets statistiquement détectables sur les

[14] Le livret personnel de compétences (LPC) a été introduit à la fin de l'année scolaire 2010-2011. Le Livret scolaire unique (LSU) lui a succédé en 2017.

[15] Depuis ce rapport de 2017, une liste de logiciel interopérable est définie sur le site eduscol et des repères annuels de progression sont identifiés pour chaque cycle en mathématiques, français et enseignement civique et moral.

apprentissages tels que mesurés par les épreuves obligatoires du DNB en fin de collège. Les élèves, qu'ils aient été exposés à l'évaluation par compétence seulement en 6è ou pendant tout ou presque leur scolarité au collège, ont obtenu les mêmes résultats au DNB que les élèves exposés au mode d'évaluation classique par notes chiffrées. Cette absence d'effets sur les résultats au DNB se retrouve chez tous les élèves quelle que soit leur origine sociale. Si quelque chose se passe, il s'agirait d'une légère baisse du taux de non-admis chez les élèves défavorisés, mais ceci reste davantage une piste à approfondir qu'une certitude » (Algan, Constantin, Delpeuch, Huillery, Prost, 2018, p.14). Les auteurs considèrent que ces résultats sont à manier avec prudence dans la mesure où il s'agit d'une évolution radicale dans les pratiques des professeurs qui va bien au-delà d'un simple changement de métrique. Une telle transformation peut nécessiter un temps de formation et d'adaptation

Jugement scolaire et biais évaluatifs

De nombreux travaux et études révèlent l'existence de normes d'évaluation implicites et de biais évaluatifs dans la production d'un jugement scolaire, qu'il s'agisse de la validation d'une compétence ou de l'attribution d'une note.

Des normes d'évaluation implicites

Les professeurs sont soumis à des normes de notation implicites qui les conduisent le plus souvent à satisfaire à une « constante macabre » (Antibi, 2003) en contrôlant leur moyenne de classe à un niveau qu'ils jugent acceptables pour ne pas être considérés comme laxistes ou trop exigeants. Ce faisant, ils ont tendance à maintenir inconsciemment un certain pourcentage d'élèves en dessous de la moyenne ce qui peut conduire à l'échec scolaire.

L'existence d'une norme d'évaluation est confirmée par une étude récente de la notation au collège (Bressoux, Lima, Rossignol, 2018). L'analyse de la variance des notes au contrôle continu entre élèves et entre établissements fait apparaître de faibles différences entre les collèges. Toutefois, la variance inter-établissement devient beaucoup plus importante lorsqu'on l'évalue en prenant en considération non plus les notes du contrôle continu mais celles du DNB. « Autrement dit, globalement, les notes varient assez peu d'un collège à l'autre mais, lorsqu'on peut prend en compte le niveau « standard » des élèves, ces différences deviennent très importantes » (p.31-32) et la prise en compte de variables individuelles (notes au DNB, sexe, boursier ou non, etc.) n'affecte pas ce constat.

Pourtant, les différences liées à l'origine sociale des élèves accueillis entre les collèges devraient induire des différences importantes dans les résultats au contrôle continu : les collèges accueillant des élèves favorisés socialement ayant en moyenne de meilleurs résultats. Or, ce n'est pas le cas. Les auteurs de l'étude avancent l'explication suivante : en l'absence de points de comparaison objectifs, les enseignants ne peuvent savoir avec précision le niveau de leurs élèves par rapport à ceux d'autres collèges. Ils se réfèrent donc à une norme d'évaluation tacite qui les incite à ne pas avoir des notes très faibles ou très élevées, qui sont toujours plus difficiles à expliquer et plus sujettes à réprobations (degré d'exigence trop élevé *versus* laxisme). C'est pourquoi, « les notes sont beaucoup plus resserrées d'un

établissement à l'autre que ne l'est le niveau "réel" des élèves » (Ibid., p.32), lorsqu'il est apprécié à partir d'un examen externe.

La diversité des biais évaluatifs

Dès les années 1960, Pierre Bourdieu et Jean-Claude Passeron ont montré que le jugement professoral était marqué par des biais de classe fondés sur des critères inconscients de perception sociale (Bourdieu, Passeron, 1964, 1970).

Depuis de nombreux travaux ont montré que tout jugement scolaire est soumis à des biais évaluatifs qui sont principalement de deux ordres (Merle, 2018) :

- ceux liés intrinsèquement à toute pratique évaluative. Il s'agit de biais cognitifs qui affectent inconsciemment celui qui juge. Par exemple, l'effet de halo illustre le fait que tout évaluateur juge inconsciemment en fonction d'un arbitraire culturel (normes, valeurs, sélections d'informations extérieures) qui affecte son appréciation en dehors des caractéristiques propres de ce qui est jugé. D'autres biais cognitifs peuvent être liés à des effets de flou liés à l'existence de critères d'évaluation implicites et à la difficulté à définir ce qui est réellement attendu ou bien encore à des effets d'ordre dans la correction des copies, etc. ;
- ceux liés aux caractéristiques des élèves, qu'elles soient scolaires (niveau, progrès réalisés, type d'établissement, niveau de la classe, etc.) ou sociales (genre, origines sociale et ethnique, etc.) (voir notamment Duru-Bellat, Mingat, 1993 ; Bressoux, Pansu 2004 ; Terrier, 2014 ; Daussin, Rocher, Troseille, 2010).

Au-delà de ces biais évaluatifs, Pierre Merle montre que l'évaluation scolaire n'est jamais mécanique et qu'elle ne reflète pas l'application stricte d'un barème. Il faut envisager l'évaluation comme le produit d'interactions à partir desquelles se construisent des arrangements évaluatifs relatifs à des « situations spécifiques, propres à la classe, à l'établissement, et aux relations particulières maître-élèves considérées individuellement et collectivement » (Merle, 2018, p.233).

Ainsi, « bien qu'étant fondé pour une large part sur les performances scolaires réelles, ce jugement [évaluatif] est sous l'influence de facteurs qui peuvent ne pas avoir de lien direct avec les compétences scolaires des élèves […], il s'élabore à partir d'éléments sociocognitifs, contextuels, et sociétaux indépendants des élèves eux-mêmes et qui en relativisent la portée. » (Bressoux, Dompnier, Pansu, 2011, p.83). Pourtant ce jugement scolaire a un impact considérable pour les élèves tant en termes de représentation de soi (Merle, 2012 ; Butera, Buchs, Darnon, 2011) que d'orientation et de parcours scolaire.

4. L'évaluation par les examens

En reposant sur des sujets et des épreuves identiques, en garantissant l'anonymat social et scolaire des candidats, les examens nationaux (DNB, baccalauréat, BTS) sont *a priori* moins soumis aux biais évaluatifs. En cela, ils peuvent constituer une mesure fiable des « acquisitions réelles » (CNESCO, 2016, p.2) des élèves et être un moyen de réduire les inégalités sociales à l'école en proposant une référence externe sur laquelle tous les enseignants, quels que soient leurs établissements d'exercice, peuvent prendre appui.

En outre, « les examens conduisant à la délivrance d'un diplôme reconnu par l'État sont organisés selon les principes de neutralité, de probité, de confidentialité et d'égalité de traitement des candidats, de la conception des épreuves à la correction des copies et l'attribution des notes. » (Cour des comptes, 2017, p.66). Ces principes sont autant de garanties d'une évaluation des acquis des élèves équitable et moins soumise aux aléas d'un contrôle continu peu organisé. Ainsi, sous réserve que l'examen couvre un champ large de disciplines et repose sur des épreuves diversifiées et suffisamment complexes pour éviter les écueils du *teaching to the test*, les examens permettent d'objectiver l'évaluation des acquis des élèves.

Toutefois, de nombreux travaux et rapports soulignent la lourdeur organisationnelle des examens, en particulier du baccalauréat, et leur coût financier, pour une portée évaluative sujette à caution.

À côté des contraintes organisationnelles (commission d'élaboration des sujets, convocation des candidats et des correcteurs, mise à disposition de locaux, organisation des corrections et des jurys, commissions d'entente et d'harmonisation, organisation des épreuves de rattrapage, etc.), des effets en termes de perte d'heures d'enseignement liées à la banalisation d'une grande partie du mois de juin (absence d'enseignements pour un grand nombre d'élèves ne passant pourtant pas d'examen et d'enseignants ne corrigeant pas d'épreuve), les examens ont un coût financier « mal connu » (Cour des comptes, 2017, p.68). Le coût du seul baccalauréat aurait été, en excluant les rémunérations des enseignants, de l'ordre de 90 à 100 millions d'euros selon le rapport des inspections générales IGF, IGAENR, IGEN et CGIET de 2011.

Parallèlement, les examens ne sont pas exempts de critiques sur les plans pédagogique et docimologique :

- les sujets sont régulièrement l'objet de remarques sur leur pertinence, leur degré de difficulté ou bien parce qu'ils comportent des erreurs ;
- le bac n'est pas une loterie (Merle, 2018, p.119-120), notamment parce qu'il comporte plusieurs épreuves, cependant la notation n'est pas exempte de biais et de distorsions d'évaluation (Suchaut, 2008). Le brassage des copies entre établissements différents, l'existence de commission de barème et de réunion d'entente et d'harmonisation et le suivi des corrections par les corps d'inspection permettent de limiter ces biais, mais ces modalités sont mises en œuvre de façon très inégale selon les académies et les disciplines (rapport IGAENR-IGEN, novembre 2017).

Enfin, la capacité des examens à mesurer réellement les acquis des élèves est contestée. Les copies d'examen ne font pas l'objet d'une analyse systématique en termes d'acquis des élèves. La variété des sujets et la variation des corrections d'une session à l'autre rend incertaine la comparaison des

résultats dans le temps. Si la DEPP a pu procéder à des études docimologiques (DEPP, décembre 2005), ces études restent rares.

En constatant d'une part l'absence de corrélation entre les taux de réussite des académies aux différents examens avec leur situation socio-économique (alors même que les résultats scolaires sont, particulièrement en France, fortement déterminés socialement) et d'autre part la déconnexion entre la stabilité des résultats aux examens nationaux et l'évolution des résultats des élèves français aux évaluations internationales, la Cour des comptes considère que la certification d'un niveau de connaissances par les résultats au baccalauréat ou au DNB ne peut être appréhendée « *qu'avec précaution* » (Cour des comptes, 2017, p.67).

Bibliographie

ALGAN Yann, HUILLERY Élise, PROST Corinne, « Confiance, coopération et autonomie : pour une école du 21e siècle », *Notes du Conseil d'analyse économique*, n°48, 2018.

ALGAN Yann, CONSTANTIN Jean, DELPEUCH Samuel, HUILLERY Élise, PROST Corinne, « Impact de l'évaluation par compétence », *Conseil d'analyse économique, Focus*, n°027-2018, octobre 2018.

ANTIBI André, *La constante macabre ou comment a-t-on découragé des générations d'élèves ?,* Paris : Nathan, 2003.

Assemblée nationale, « L'organisation de la fonction évaluative du système éducatif », *Rapport d'information de l'Assemblée nationale*, n°1265, septembre 2018.

BAUDELOT Christian, ESTABLET Roger, L'élitisme républicain. *L'école française à l'épreuve des comparaisons internationales,* Paris : Seuil (« La République des idées »), 2009.

BOURDIEU Pierre, PASSERON Jean-Claude, *Les Héritiers*, Paris Minuit, 1964.

BOURDIEU Pierre, PASSERON Jean-Claude, *La Reproduction*, Paris : Minuit, 1970.

BORST Grégoire, HOUDÉ Olivier, *Le Cerveau et les apprentissages*, Paris : Nathan (« Repères pédagogiques »), 2021.

BRAXMEYER Nicole, GUILLAUME Jean-Claude, LÉVY Jean-François, « Les pratiques d'évaluation des enseignants au collège », *Note d'évaluation*, n°04.13, DEPP, 2004.

BRESSOUX Pascal, PANSU Pascal, *Quand les enseignants jugent leurs élèves*, Paris : PUF, 2003.

BRESSOUX Pascal, DOMPNIER Benoît, PANSU Pascal, « L'évaluation scolaire : une activité multidéterminée » *in* Fabrizio Butera *et al*., *L'évaluation, une menace ?*, Paris : PUF, 2011.

Centre national d'étude des systèmes scolaires (CNESCO) :

- *L'évaluation des élèves par les enseignants dans la classe et dans les établissements : réglementation et pratiques. Une comparaison internationale*, décembre 2014.
- *Les grands débats du baccalauréat*, juin 2016.
- *L'orientation en fin de 3ème : déterminants individuels et contextuels*, décembre 2018.

Conseil scientifique de l'éducation nationale, *Évaluer pour mieux aider*, Ministère de l'Éducation nationale, Réseau Canopé, mai 2019.

Comité éthique et scientifique de Parcoursup, Rapports 2019, 2020, 2021, Ministère de l'enseignement supérieur, de la recherche et de l'innovation.

Cour des comptes, « L'éducation nationale : organiser son évaluation pour améliorer sa performance », décembre 2017.

DEMAILLY Lise, « L'évaluation de l'action éducative comme apprentissage et négociation », *Revue française de pédagogie*, n°142, p.115-129, janvier-février-mars 2003.

DEROUET Jean-Louis, NORMAND Romuald, *Le développement d'une culture de l'évaluation dans l'Éducation Nationale : comment les enseignants utilisent-ils les résultats des évaluations nationales ?,* INRP, 2003.

Direction de l'évaluation, de la prospective et de la performance (DEPP) :

- « Étude docimologique du baccalauréat de la session 2003 », *Dossiers n°72*, décembre 2005.
- « Étude docimologique du baccalauréat scientifique», *Note d'information n°05-38*, décembre 2005.
- DAUSSIN Jeanne-Marie, ROCHER Thierry, TROSEILLE Bruno, « L'attestation de la maîtrise du socle commun est-elle soluble dans le jugement des enseignants ? », *Éducation & Formation,* n°79, p.45-58, décembre 2010.
- PERRONNET Sabrina, « Temps de travail des enseignants du second degré du public : pour une heure de cours, une heure de travail en dehors », *Note d'information,* n°13.13,2013.
- ROCHER Thierry, TROSSEILLE Bruno, « Les évaluations standardisées des élèves. Perspective historique », *Éducation & Formation*, n°86-87, p.15-35, mai 2015.
- KESKPAIK Saskia, ROCHER Thierry, « La motivation des élèves français face à des évaluations à faibles enjeux », *Éducation & Formation*, n°86-87, p.119-139, mai 2015.
- État de l'école, DEPP, 2021.

DURU-BELLAT Marie, « Usages et mésusages des enquêtes PISA dans l'évaluation et le pilotage des systèmes éducatifs », *LIEPP Policy Brief*, 2012.

DURU-BELLAT Marie, MINGAT Alain, *Pour une approche analytique du fonctionnement du système éducatif*, Paris : PUF, 1993.

DUPRIEZ Vincent, MALET Régis (dir.), *L'évaluation dans les systèmes scolaires. Accommodement du travail et reconfigurations des professionnalités*, Bruxelles : De Boeck, 2013.

DUPRIEZ Vincent, MONS Nathalie, « Les politiques d'accountability », *Recherche et formation*, n°65, p.45-59, 2010.

DUTERCQ Yves, LANÉELLE Xavière, « La dispute autour des évaluations des élèves dans l'enseignement français du premier degré », *Sociologie*, n°1, p.43-62, 2013.

EURYDICE, Les évaluations standardisées des élèves en Europe : objectifs, organisation et utilisation des résultats, Commission européenne, 2009.

FELOUZIS Georges, CHARMILLOT Samuel, *Les Enquêtes PISA*, Paris : PUF (« Que sais-je ? »), 2012.

HADJI Charles, *Faut-il avoir peur de l'évaluation ?*, Louvain-la-Neuve : De Boeck, 2012.

Haut conseil à l'évaluation de l'école, *Apprécier et certifier les acquis des élèves en fin de collège : diplôme et évaluations-bilans, Avis du HCéé*, n°2, juin 2001.

Haut conseil à l'éducation, *Les indicateurs relatifs aux acquis des élèves, Bilan des résultats de l'École*, 2011.

GRENET Julien, « PISA : une enquête bancale », *La Vie des Idées*, février 2008.

Rapports des inspections générales du MEN et du MESRI :

- *Les acquis des élèves, pierre de touche de la valeur de l'école ?*, n°2005-079, juillet 2005.
- *Proposition pour une évolution du baccalauréat*, décembre 2011. (Rapport IGF, IGAENR, IGEN et CGIET)
- *Les évaluations nationales et internationales sur les acquis des élèves et sur d'autres dimensions des systèmes éducatifs*, n°2012-056, mai 2012.

- *Suivi qualitatif des conditions de passation et de correction des évaluations nationales exhaustives de CE1 et de CM2 en 2012*, n°2012-98, juillet 2012.
- *La mise en œuvre du LPC au collège*, n°2012, août 2012.
- La *notation et l'évaluation des élèves éclairées par des comparaisons internationales*, Rapport IGEN n°2013-072, juillet 2013.
- *L'outil « livret scolaire unique » et son impact sur les modalités d'évaluation*, n°2017-092, octobre 2017.
- *Processus qualité et sécurisation des examens*, n°2017-046, novembre 2017.
- *L'organisation et la passation des évaluations nationales dans le premier degré*, n°2019-096, octobre 2019.
- *L'exploitation des évaluations nationales CP CE1*, n°2020-020, janvier 2020.
- *Les évaluations nationales du second degré*, n°2020-074, juin 2020.

LEVASSEUR Jacqueline, « L'évaluation nationale des acquis des élèves. Finalités et exploitations », *Revue internationale d'éducation de Sèvres*, n°11, p.101-114, 1996.

MAROY Christian, « Vers une régulation post-bureaucratique des systèmes d'enseignement en Europe ? », *Sociologie et société*, volume 40, n°1, p.31-55, 2008.

MERLE, Pierre, *L'élève humilié. L'école, un espace de non-droit ?*, Paris : PUF, 2012.

MERLE Pierre, *Les pratiques d'évaluation scolaire*, Paris : PUF, 2018.

MEURET Denis, « États-Unis : quelle suite pour "No Child Left Behind" ? », *Revue internationale d'éducation de Sèvres*, n°69, p.27-30, septembre 2015.

Ministère de l'Éducation nationale, de l'enseignement supérieur et de la recherche :

- Conférence nationale sur l'évaluation des élèves, 2014.
- Évaluer dans le cadre du contrôle continu. Baccalauréat 2021, terminales générales et technologiques, 2021.
- Guide pour l'évaluation des apprentissages et des acquis des élèves dans le cadre de la réforme du lycée général et technologique, septembre 2021.

MOTTIER LOPEZ Lucie, CRAHAY Marcel, *Évaluations en tension. Entre la régulation des apprentissages et le pilotage des systèmes*, Louvain-la-Neuve : De Boeck Supérieur, 2009.

MONS Nathalie, « Effets théoriques et réels des politiques d'évaluation standardisée », *Revue française de pédagogie*, n°169, p.99-140, octobre-décembre 2009.

MONS Nathalie, PONS Xavier, « Pourquoi n'y a-t-il pas eu de « choc PISA » en France ? Sociologie de la réception d'une enquête internationale (2001-2008) », *Revue française de pédagogie*, n°182, p.9-18, 2013.

OCDE, *Synergies for better learning*, 2013. Résumé en français.

OCDE, *Education Policy Outlook 2019* (chapter 3. "Evaluation and assessment: Policy priorities and trends", 2008.

PONS Xavier, « Quarante ans d'évaluation ministérielle des acquis des élèves en France : complexification et politisation », *Politiques sociales et familiales*, n°110, p.9-18, 2012.

Revue française de pédagogie, « PISA : analyses secondaires, questions et débats théoriques et méthodologiques », volume 4, n°157, 2006.

REY Olivier, FEYTANT Annie, *Évaluer pour (mieux) faire apprendre*, *Dossier de veille de l'Ifé*, n°94, septembre 2014.

ROZENWAJN Esteban et DUMAY Xavier, « Les évaluations externes à faibles enjeux : les enseignants confrontés à eux-mêmes ? », *Revue française de pédagogie*, vol. 194, n°1, p.71-90, 2016.

SUCHAUT Bruno, « La loterie des notes au bac. Un réexamen de l'arbitraire de la notation des élèves », *Documents de travail*, IREDU, 2008.

THÉLOT Claude, « Évaluer l'École », *Études*, Tome 397, p.323-334, octobre 2002.

TERRIER Camille, « Un coup de pouce pour les filles ? Les biais de genre dans les notes des enseignants et leur effet sur le progrès des élèves », *Les notes de l'IPP*, n°14, décembre 2014.

Liste des sigles utilisés

AFNOR : Association française de normalisation

CAP : Certificat d'aptitude professionnelle

CP : Cours préparatoire

CE1 / CE2 : Cours élémentaire première année / cours élémentaire seconde année

CEDRE : Cycle des évaluations disciplinaires réalisées sur échantillons qui évalue les acquis des élèves de CM2 et 3ème.

CNESCO : Centre national d'étude des systèmes scolaires

CM1 / CM2 : Cours moyen première année / cours moyen seconde année

CSEN : Conseil scientifique de l'éducation nationale

COFO : Évaluation des compétences fondamentales réalisée en Suisse

DEPP : Direction de l'évaluation, de la prospective et de la performance du ministère de l'éducation nationale, de la jeunesse et des sports

DGESCO : Direction générale de l'enseignement scolaire du ministère de l'éducation nationale, de la jeunesse et des sports

DNB : Diplôme national du brevet

HCéé : Haut conseil à l'évaluation de l'école

ICCS : *International Civic and Citizenship Education Study.* Étude comparative internationale conçue par l'IEA qui aborde les connaissances, les attitudes et l'engagement civiques et citoyens des élèves scolarisés au huitième grade (4ème pour la France).

ICILS : *International Computer and Information Litteracy Study*. Étude organisée par l'IEA pour évaluer les performances des élèves de quatrième en littératie numérique ainsi qu'en pensée informatique.

IEA : *International Association for the Evaluation of Educational Achievement*. Association internationale d'instituts de recherche nationaux, d'organismes de recherche publique, d'universitaires et d'analystes qui mesure les résultats des élèves notamment dans des matières telles que les mathématiques et les sciences (enquêtes TIMSS), la lecture (enquêtes PIRLS).

IGAENR : Inspection générale de l'administration, de l'éducation nationale et de la recherche

IGEN : Inspection générale de l'éducation nationale

IGÉSR : Inspection générale de l'éducation, du sport et de la recherche. Inspection créée en 2019, elle fusionne plusieurs inspections dont l'IGAENR et l'IGEN.

INSPE : Institut national supérieur du professorat et de l'éducation

IH2EF : Institut des hautes études de l'éducation et de la formation

LOLF : Loi organique relative aux lois de finance

OCDE : Organisation de coopération et de développement économiques

PIRLS : *Progress in International Reading Literacy*. Étude comparative internationale conçue par l'IAE qui mesure les performances en compréhension de l'écrit des élèves en fin de quatrième année de scolarité obligatoire (CM1 pour la France).

PISA : Programme international pour le suivi des acquis des élèves. Évaluation créée par l'OCDE qui vise à tester les compétences des élèves de 15 ans en lecture, sciences et mathématiques. Cette évaluation se déroule tous les 3 ans.

TIMSS : *Trends in Mathematics and Science Study*. Étude comparative internationale conçue par l'IEA qui mesure le niveau des connaissances scolaires des élèves à la fin de la quatrième année de scolarité obligatoire (CM1 pour la France) et de huitième grade (4ème pour la France) en mathématiques et en sciences.

UNESCO : Organisation des Nations unies pour l'éducation, la science et la culture

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Conseil d'évaluation de l'École

Le Conseil d'évaluation de l'École est chargé d'évaluer en toute indépendance le système scolaire, son organisation et ses résultats. Il produit des rapports, avis et recommandations visant à l'amélioration du service public de l'éducation.
Les délibération et résultats des travaux du Conseil sont publics, ils ont vocation à éclairer les pouvoirs publics et enrichir le débat sur l'éducation.

Cette note a été conçue collégialement, sous la direction des membres du Conseil, avec des chercheurs et scientifiques disposant d'une expertise sur ces questions, ainsi que des inspecteurs généraux et des spécialistes de l'évaluation ou des statistiques, les directions du ministère et des acteurs des académies.

Avertissement :
Ce document est dans le domaine public. L'autorisation de le reproduire en tout ou en partie est accordée, toutefois cette reproduction doit :

- privilégier les citations in extenso afin de ne pas modifier le sens du texte. Si un changement s'avère nécessaire il ne doit concerner que la forme ou des éléments secondaires de la citation ;
- mettre en évidence les parties relevant du CEE de celles relevant de la nouvelle publication ;
- toujours être créditée : Conseil d'évaluation de l'École, Renforcer la cohérence des évaluations des acquis des élèves : un enjeu de réussite et d'équité, Document de travail 22-02, [Février 2022].

Retrouvez toute la documentation et les publications du CEE sur :
education.gouv.fr/CEE

Conseil d'évaluation de l'École
61-65, rue Dutot – 75732 Paris cedex 15

cee@cee.gouv.fr

Conception graphique : opixido.com